DOCUMENT RESUME

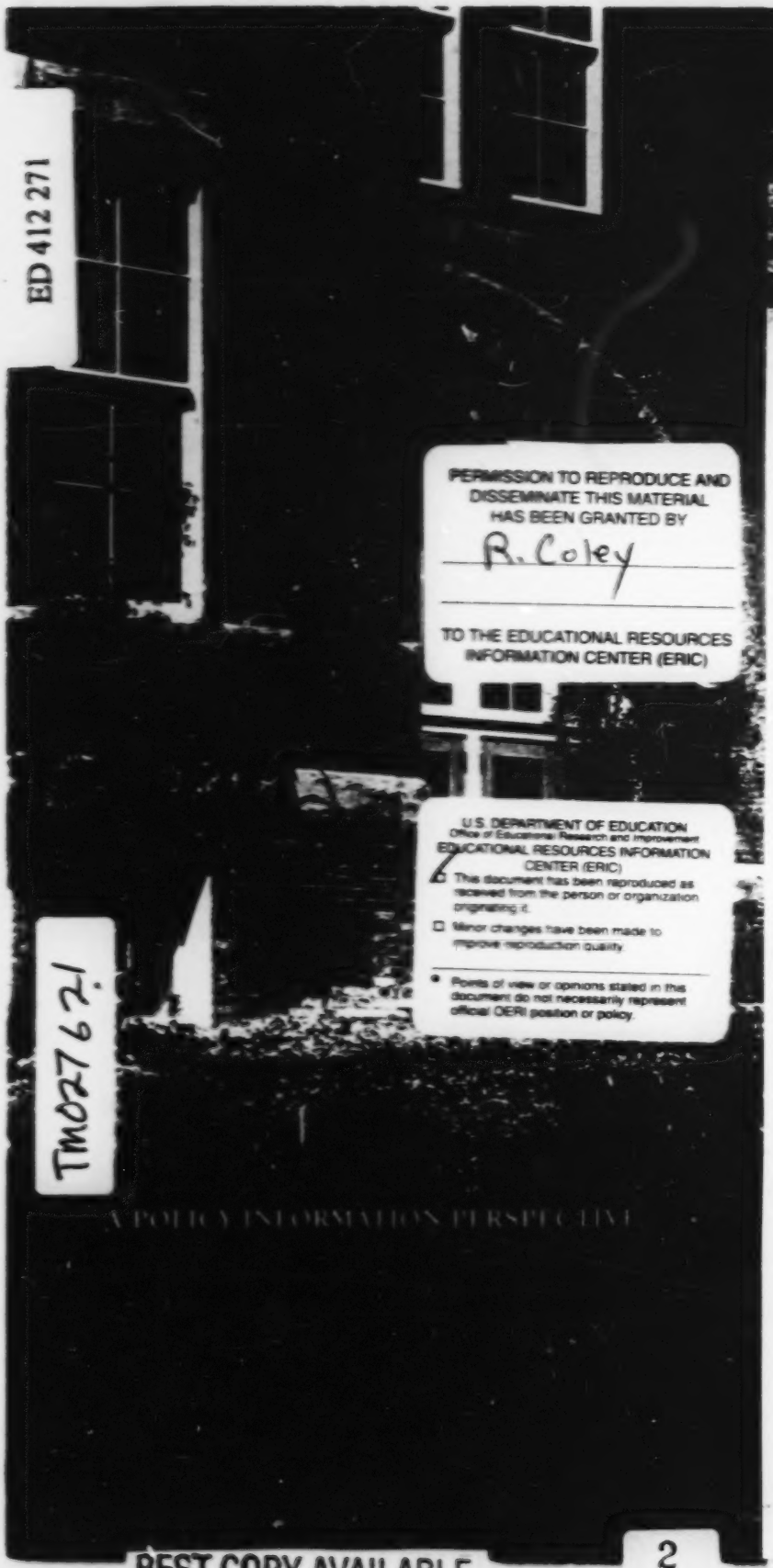
ED 412 271 TM 027 621

AUTHOR Wenglinsky, Harold
TITLE When Money Matters: How Educational Expenditures Improve Student Performance and How They Don't. A Policy Information Perspective.
INSTITUTION Educational Testing Service, Princeton, NJ. Policy Information Center.
SPONS AGENCY National Science Foundation, Arlington, VA.
PUB DATE 1997-04-00
NOTE 53p.
CONTRACT REC-9628157
AVAILABLE FROM Policy Information Center, Mail Stop 04-R, Educational Testing Service, Rosedale Road, Princeton, NJ 08541-0001; internet: http://www.ets.org ($9.50).
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; Behavior Problems; Databases; Educational Administration; Educational Environment; *Educational Finance; *Elementary Secondary Education; *Expenditures; Financial Support; Grade 4; Grade 8; Multivariate Analysis; National Surveys; *Performance Factors; *Resource Allocation; Social Influences; Teacher Education; Teacher Student Ratio
IDENTIFIERS Common Core of Data Program; *National Assessment of Educational Progress

ABSTRACT

Little agreement exists on which school expenditures and resources are most likely to improve student resources or whether resources really matter at all. This study compiles a national database of school finance information and analyzes the data to address the importance of school expenditures. Data were collected from the National Assessment of Educational Progress, a nationally representative sample of fourth and eighth graders, the Common Core of Data of the U.S. Department of Education, and a Teacher's Cost Index, which measures variations in the costs of education. Data were analyzed with multivariate techniques to produce flow charts for grades four and eight of how dollars and resources influence student achievement in mathematics. The study found that expenditures affect the achievement of fourth graders in two steps and eighth graders in three. For fourth graders, the process begins with increased expenditures on instruction and school district administration which then increase teacher-student ratios. The increased teacher-student ratios raise average achievement in mathematics. For eighth graders, the first step is the same. Then increased teacher-student ratios reduce problem behaviors and improve the social environment of the school. A lack of problem behaviors and a positive social environment then increase average achievement in mathematics. Expenditures related to capital outlays, school level administration, and teacher education levels were not found to increase achievement. Findings support both a productivity perspective and a traditional perspective on school finance. An appendix discusses the research methodology. (Contains 3 figures, 3 tables, and 41 references.) (SLD)

ED 412 271



PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
R. Coley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM027621

A POLICY INFORMATION PERSPECTIVE

How

EDUCATIONAL EXPENDITURES IMPROVE STUDENT PERFORMANCE AND HOW THEY DON'T

POLICY INFORMATION CENTER
Educational Testing Service
Princeton, New Jersey 08541-0001

BEST COPY AVAILABLE

Additional copies of this report can be ordered for $9.50 (prepaid) from:

Policy Information Center
Mail Stop 04-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541-0001
(609) 734-5694
Internet - pic@ets.org
http://www.ets.org

Copyright © 1997 by Educational Testing Service. All rights reserved. Educational Testing Service is an Affirmative Action/ Equal Opportunity Employer.

Educational Testing Service, *ETS*, and ® are registered trademarks of Educational Testing Service.

April 1997

TABLE OF CONTENTS

Preface iii

Acknowledgments v

Executive Summary vii

Introduction 1

Policy Background 6

Prior Research on the Spending-Achievement Relationship 10

Data Sources and Study Method 14

Findings 18

- Description of School Districts 18
- Testing Relationships 21
- Comparing Subgroups 23
- Interpretations 26
- Knowing More 28

Policy Implications 30

Appendix: How the Study was Conducted 32

- Data 32
- Method 36

Variable Definitions 41

References 42

PREFACE

Since the publication of Equality of Educational Opportunity by James S. Coleman in 1965, there has been an enduring discussion of whether differences in the expenditure of resources on the schools makes any difference in educational achievement. No consensus has yet been reached, despite the fact that some of our best academic minds have focused on the question.

The Coleman Report, as it became known, was based on extensive national tests and surveys. Research since then has been localized, often without direct measures of student achievement, and with only a gross measure of resources such as per capita student expenditures. This report, authored by Harold Wenglinsky, associate research scientist at the ETS Policy Information Center, attempts to take the scale of research back to the national level of the Coleman report, and beyond it in terms of the depth of the data and the incorporation of decades of "effective schools" research. This policy report draws on two technical reports written by Wenglinsky at ETS as a NAEP Fellow, with partial funding from the National Science Foundation. They are available as Research Reports from ETS.

Wenglinsky draws on the respected National Assessment of Educational Progress (NAEP) for achievement data and for data on the classroom context and home backgrounds of the students. Instead of using only the gross measure of expenditures used in previous research, he has married the NAEP data base and the Common Core data base at the District level, enabling him to trace the effects of different types of expenditures, such as for instruction, for the district office, for the principal's office, and for capital outlays, on the schools. Also, he has adjusted for cost of differences by using a teacher cost index produced by the National Center for Education Statistics.

Wenglinsky describes the unfolding saga of the legal battles over school finance, and the results of past research, as well as presenting his own empirical findings. Also, he offers his judgment as to the policy implications of these findings and the need for further investigation. This report is issued in the Center's Policy Issue Perspectives series, where we encourage both research and professional judgment as to what the research means.

Whether—and how—"money matters" is a most controversial issue, with debate polarized in legal, political, and academic settings. In the academic sphere, disagreement abounds, even on what past research shows. Debate on this issue will not soon come to closure. And until it does, well-informed debate is a healthy way to proceed. We hope this report makes a contribution.

Paul E. Barton
Director
Policy Information Center

ACKNOWLEDGMENTS

This report could not have been written without the contributions of many individuals and institutions. Financial support for the research and technical reports on which this policy report is based was provided in part by the National Science Foundation, under grant REC-9628157, and in part by the Educational Testing Service.

Eugene Johnson and Paul Barton provided invaluable guidance and feedback throughout the course of the study and the writing of this and other reports. Alfred Rogers assisted in putting together the National Assessment of Educational Progress (NAEP) data base that I used in my study. Keith Rust assisted in linking NAEP to the Common Core of Data. William Fowler provided guidance on how to adjust per pupil expenditures for variations in the cost of education. Russell Almond provided excellent comments on the technical report for this study. Margaret Goertz and Allan Odden provided excellent comments on an earlier draft of this report. I also benefited from the input, at various times, of Henry Braun, Richard Coley, Charles Davis, Jeremy Finn, Jan-Eric Gustafsson, Robert Mislevy, Eiji Muraki, James Roberts, Juliet Shaffer, Richard Snow and Larry Suter. Carla Cooper provided desktop publishing services, Patricia Ciaccio did the editing, and Kelly Gibson was the designer. Any errors of fact or interpretation included in this report, however, are the responsibility of the author.

Executive Summary

Significant inequalities in school expenditures and resources remain, even after 30 years of court decisions designed to reduce these inequalities. Nearly 50 percent of the funding for U.S. schools comes from property taxes levied in school districts.

Because the amount of taxable wealth varies greatly from district to district, this results in large variations in the amount of money school districts have to spend. Thus, students from the poorest neighborhoods are more likely to attend schools that lack important resources. The gaps in spending between the districts of the wealthy and those of the poor have been reduced somewhat as a result of litigation, beginning with the case of *Serrano v. Priest* in 1971. State courts and legislatures have provided funds to reduce disparities between school districts and to raise aggregate spending in all school districts in the state. Nevertheless, significant disparities remain because of the continued dependence of school districts on property taxes, the limited degree to which states are willing to provide funds to reduce disparities, and the far more limited role the federal government is willing to play.

Policymakers are divided in their views on the proper course to follow in school finance. Some argue for continuing the traditional approach to school finance reform. They feel that more money needs to be spent to reduce disparities between rich and poor school districts to the point where spending levels in the two types of district are equivalent. Some even suggest raising spending levels in poor school districts above those in affluent ones to compensate for other inequalities that students in poor districts experience. Other policymakers argue for a "productivity" approach to school finance reform. They note that significant increases in spending and the reduction of inequalities in spending have not netted the large increases expected in achievement. Therefore, they call for reallocating existing funds and earmarking new funds to those areas most likely to improve teaching and learning. In their view, most of the ways in which the additional dollars made available to districts are conventionally spent do not contribute to improvements in student achievement. Consequently, new and innovative approaches for linking dollars to achievement must be developed and tested.

Unfortunately, the research base for these two competing views is limited. Little agreement exists on which expenditures and resources are most likely to improve student performance or on whether resources matter at all. The Colemen Report, a study conducted in 1966, opened the debate. The Coleman Report held that resources made little difference to achievement once the background characteristics of students were taken into account. The debate has continued over the past 30 years, with some studies finding a relationship and others not finding one. This stalemate is due both to a lack of research on the scale of the original Coleman Report and to some methodological problems in the studies.

The current study seeks to begin to break this deadlock, through compilation of a national data base of school finance information and statistical analysis of the data base to address methodological issues in previous research. Data were collected from three sources: (1) the National Assessment of Educational Progress, a nationally representative sample of fourth and eighth graders who took achievement examinations in mathematics and were asked questions pertaining to their background characteristics and the climate of the school; (2) the Common Core of Data, a data base of school finance information collected by the U.S. Department of Education from all school districts in the nation; and (3) the Teacher's Cost Index, also developed by the U.S. Department of Education, which measures variations in the cost of education between states. The data were analyzed using advanced multivariate techniques in order to produce flow charts for fourth and eighth graders of how dollars and resources influence student achievement in mathematics.

The study found that expenditures can affect the achievement of fourth graders in two steps and of eighth graders in three steps. For fourth graders, the process is as follows:

Step 1: Increased expenditures on instruction and school district administration increase teacher-student ratios.

Step 2: Increased teacher-student ratios raise average achievement in mathematics.

For eighth graders, the process is:

Step 1: Increased expenditures on instruction and school district administration increase teacher-student ratios.

Step 2: Increased teacher-student ratios reduce problem behaviors and improve the social environment of the school.

Step 3: A lack of problem behaviors among students and a positive social environment raise average achievement in mathematics.

In addition, the study found that variations in other expenditures and resources were not associated with variations in achievement. Variations found not to be related in this way were:

1. Capital outlays (spending on facility construction and maintenance)
2. School level (principal's office) administration
3. Teacher education levels

This study provides some support for the productivity perspective and some support for the traditional perspective on school finance. It supports the notion among productivity researchers and policymakers that some dollars matter more than others. It also finds, however, that some traditional spending practices of school districts (spending for teacher-student ratios and central office administration) are conducive to academic achievement. On the basis of the findings, it is suggested that courts and legislatures, in raising additional funds for school districts, concern themselves with productivity. In earmarking funds and identifying priorities, however, they should make sure to include the conventional inputs found here to be important.

INTRODUCTION

The Newark school system reflects within it all of the tensions of school finance. Newark is a city made up overwhelmingly of poor, minority families. Its school system, reflecting a weak tax base, spent well below the state average.

After a series of State Supreme Court decisions, however, spending in the system was increased, to $8,829 per pupil in 1994 ($1,327 above the state average for that year). Nevertheless, student achievement was still at worrisomely low levels. Among 11th graders, for instance, the percentages of students passing proficiency tests in 1994 were 42 percent for reading, 57 percent for writing, and 41 percent for mathematics, compared with state averages of 83 percent, 89 percent, and 84 percent, respectively. The state subsequently took over the Newark school system and hired a new superintendent, Dr. Beverly Hall, to make major changes. She found that lack of resources, while a real problem, was not the only problem in the system. One problem was the way money was spent. School janitors were making 36 percent more than the regional average, although they cleaned 20 percent less building space. Bus attendants, whose job was to watch students on school buses, were paid $27,000 for three hours of work a day. Staffing patterns in the school cafeteria produced an average cost of $4.00 per lunch, compared with the national average of $2.00. In July of 1996, Dr. Hall found that she could lay off more than 500 employees; this began a process of streamlining school expenditures, with the goal of raising student achievement (Petersen 1996).

This story exemplifies the key division among advocates of school finance reform. School finance reform traditionally has meant the reduction in aggregate differences in spending between the school districts of rich students and those of poor students. This traditional view has pointed out that vast inequalities in spending exist, both between school districts within states and between states. In New Jersey in 1990, for instance, the highest-spending district spent $8,462 per pupil, compared with $5,162 in the lowest-spending district (General Accounting Office 1995). The traditional approach has been to reduce these inequalities, either by putting political pressure on state legislatures to provide extra funding to low-spending school districts or by suing for redress in state courts. The result has been legislation or court orders in nearly every state to reduce district inequalities. One of these court decisions, in New Jersey, led the state to pass the Quality Education Act, which

increased per pupil expenditures in Newark.[1]

The more recent productivity perspective on school finance reform has emphasized the need for low-achieving school districts to cut wasteful spending and to invest resources in those areas most conducive to raising achievement. In this view, equalizing spending between school districts will not by itself improve resource-poor school districts. In addition, the spending practices of resource-poor school districts must be modified. Less money should be spent on the district-level bureaucracy, and school principals should have more budgetary discretion. Support staff, as well as teachers, should not be paid out of proportion to what the market will bear. The state should earmark money for cutting-edge initiatives (such as performance-based compensation for teachers whose students show achievement increases and professional development for teachers who want to learn new skills). Ideally, the school district should cease to exist as an intermediary between the state and the school. The state should set standards for schools and provide them funds with some guidelines; the schools, governed by their principals and collaborative councils of teachers and parents, should make the more specific operational and budgetary decisions.[2] Many reform-minded superintendents have indeed sought to link their spending practices to improvements in teaching and learning; even when they have obtained additional resources from grants or state equalization funds, they have tried to spend dollars wisely.[3]

These two perspectives, one emphasizing aggregate equalization and the other emphasizing productivity, both make certain empirical assumptions about the relationship between school finance systems and the quality of education. The traditional view holds that nearly all forms of educational spending have an effect on student performance; therefore, additional funds for all of these forms will raise student achievement.[4] Spending on school facilities has an effect because dilapidated facilities show students that the system does not care about them, whereas extensive and well-maintained facilities provide an encouraging physical environment in which students can learn.

[1] For a recent example of this view on school finance reform, see Kozol (1991). The book that served as the intellectual basis for most of the court decisions and legislative initiatives equalizing education funding was Coons, Clune and Sugarman (1970).

[2] A leading exponent of this view is Allan Odden. See Odden (1993) for his view of recent court decisions and Odden and Clune (1995) for a discussion of educational productivity. See also Hanushek (1994) and Fuhrman (1996).

[3] A third perspective on school finance reform is to reduce investment in public schools and use the money to provide students with vouchers to attend private schools. This perspective is based on the empirical claim that students perform better in private schools (a claim outside the scope of this study to evaluate). For a recent study on this question, see Wenglinsky (1996a).

BEST COPY AVAILABLE

Spending on the administrative apparatus of a school and school district is worthwhile because it permits the school system to allocate resources effectively, direct instructional policy, and provide support services (such as guidance counseling and transportation) to students. Spending on teacher salaries has an effect because the quality of teaching is related to these salaries. Higher salaries allow schools to recruit better teachers, particularly those with better educational credentials or more experience; this, in turn, raises student achievement. Finally, spending to increase teacher-student ratios has an effect on students. Students in large and overcrowded classes receive less individual attention from teachers, have less opportunity to participate in class, and have less opportunity to create a cohesive social group with one another; in turn, all of these reduce student achievement. Because all of these factors play a role in student performance, the traditional view argues, the appropriate strategy to provide equal opportunity to all students is to equalize spending in all of these categories between school districts. This should be done by creating finance equalization formulas that reduce differences in aggregate per pupil expenditures, thereby providing more funds to low-performing school districts.

The productivity view, on the other hand, holds that few (if any) of the conventional forms of spending, when increased, result in achievement gains. Increases in school district administration are unlikely to raise student achievement. The superintendent's office is more typically a sink of waste, fraud, and abuse. It also often creates obstacles for principals who want to run their schools innovatively. Increasing resources for facilities is unlikely to raise student achievement, because most of the facilities in which schools invest have little to do with learning; it is unclear what the link is between a school constructing an Olympic-sized swimming pool and students showing proficiency in mathematics. Increases in teacher salaries and teacher-student ratios (the other ways in which additional resources conventionally are invested) have not seemed to net increases in student achievement. School districts, as they have been given more and more money over the past 30 years, have invested in these two areas, but there have not been comparable increases in student achievement over

* "Effect" is the standard term used in the literature to refer to a relationship in which one factor (such as educational spending) precedes and is associated with another (such as academic achievement). Since most of these studies use cross-sectional data, it is possible, although often implausible, that this assumption is untrue. In this study, the term "effect" connotes the assumed precedence of the two factors being examined; it is assumed that spending precedes achievement, rather than vice versa.

time. The establishment of a relationship between money and performance will require new and innovative fiscal practices (such as school-based budgeting and performance-based compensation) as well as the elimination of many current wasteful fiscal practices.[1]

Unfortunately, the current research base for assessing the competing claims of the traditional and productivity perspectives is inadequate to the task. The last large-scale national study of school finance, the Coleman Report, published its findings in 1966 (Coleman et al. 1966). It supported the claims of the productivity perspective that current resource inputs (including per pupil expenditures), teacher quality measures (education and experience), and teacher-student ratios did not influence student achievement. Rather, the background characteristics of the students themselves accounted for differences between rich and poor school districts. Subsequent research, generally on a state- or system-level scale, has produced contradictory results, with some studies finding a relationship between spending and achievement and other studies finding no such relationship. Studies summarizing these studies (known as "meta-analyses") also have not been able to agree on what can be concluded from prior research.

The current study, through the analysis of a large-scale national data base constructed from a variety of sources, finds that both the productivity and traditional views are correct—to a point. The study finds that, for nationally representative samples of fourth and eighth graders, variations in teacher-student ratios, expenditures on instruction, and expenditures on school district administration are positively associated with variations in mathematics achievement but that expenditures on facilities, expenditures to recruit highly educated teachers, and expenditures on school-level administration are not. Thus, the productivity researchers are correct in arguing that increasing spending in conventional areas does not result in increases in student achievement and that the key to school finance reform is for state legislatures to earmark resources for those inputs that do raise achievement. The productivity researchers, however, have rejected too quickly some of the traditional inputs (such as teacher-student ratios), increases of which do appear to raise achievement and have

[1] The traditional and productivity perspectives, as summarized here, represent pure types. Many researchers and policymakers borrow heavily from the opposite perspective. Many traditionalists, for instance, while primarily concerned with increasing aggregate resources for poor school districts, caution that this money should not be wasted (e.g., Murnane and Levy 1995). Many productivity researchers, while primarily concerned with increasing the efficiency of poor school districts, caution that the quest for efficiency should not be used as a rationale for taking money away from them (e.g., Corcoran and Goertz 1995).

BEST COPY AVAILABLE

favored some innovations (such as school-based budgeting), when some school districts may do a good job of allocating resources.

Before discussing the findings and suggesting what they imply for the two school finance perspectives, it is worthwhile to provide some background. This paper begins with a review of the evolution of policies on school finance over the past 30 years. It then summarizes prior research on school finance and suggests some shortcomings of that research that may be responsible for the lack of a consensus on the spending-achievement relationship. The paper briefly touches on the sources of data and the methodology of the study, then presents the findings and their implications.

POLICY BACKGROUND

The movement to equalize spending among school districts has arisen in response to the way in which most states finance their schools—through local property taxes. Generating revenue through local property taxes creates large inequalities among school districts, because the amount of revenue is a function of the level of wealth of the district.

As a result, school districts made up predominantly of poor families net less revenue than school districts made up predominantly of affluent families. The goal of the finance equalization movement has been to use state and federal funding to reduce the inequalities inherent in property tax funding, by providing funds to districts that have not been able to generate as much revenue from property taxes as other districts (Coons et al. 1970; Kozol 1991).

The pressure to equalize spending has been around almost as long as property taxes. In *Springfield Township v. Quick et al.* (Supreme Court, December Term, 1859:56-60), plaintiffs sued the state of Indiana for not sufficiently redressing inequalities in spending between rich and poor school districts. The case was unsuccessfully appealed to the Supreme Court. Later cases included *Stuart v. Kalamazoo* (30 Mich. 69 [1874]) and *Sawyer v. Gilmore* (109 Me. 169, 83 A. 673 [Me. 1912]). All held that the financing of school districts was at the discretion of state legislatures and that it was not within the purview of state or federal courts to interfere in this process. This remained the position of most courts through *McInnis v. Ogilvie (Shapiro)* (394 U.S. 322, 89 S. Ct. 1197 [1969]), in which the Supreme Court refused to hear a case in which the lower court had decided that resource inequalities between school districts in Illinois did not warrant court intervention.

The decision of the California Supreme Court in *Serrano v. Priest (I)* (5 Cal 3d 584, 96 Cal. Rptr. 601, 487 P.2nd 1241 [Calif. 1971]) opened the floodgates for school finance equalization. In that case, John Serrano brought suit against the state of California on the grounds that there were wide disparities in educational expenditures between school districts and that these disparities were at odds with the fundamental interest of the state to provide an education to its citizens. The court held for the plaintiff, arguing that the quality of education a student received did seem to depend upon the "resources of his [or her] school district and ultimately upon the pocketbook of his [or her] parents." This situation was at odds with the state's "fundamental interest" in education. The trial court acknowledged that the details of school finance were at the discretion of the state legislature, but it provided four ways for the legislature to comply with its constitution's "fundamental interest" in education. It could (1) abolish property taxes and fully fund schools through a statewide tax of some sort; (2) collapse existing school districts into larger ones that encompass both rich and poor populations; (3) remove commercial properties from property taxation; or (4) guarantee each school

district a given level of resources, depending on the percentage of wealth levied in property taxes (known as "power equalization").

The outcome in *Serrano* spilled over to other states. Minnesota, New Jersey, and New York all heard cases on this issue during the next year. Minnesota and New Jersey ruled in favor of greater equalization (New York did not). The trend toward school finance equalization was slowed somewhat when the U.S. Supreme Court heard the case of *San Antonio Independent School District v. Rodriguez* (411 U.S. 1, 93 S. Ct. 1278, 1973), in which the Court, while acknowledging the importance of equalizing school finances, thought it to be a concern primarily of the states and their legislatures and accordingly found for the defendant. The decision closed the door on federal litigation. At the state level, subsequent cases were decided in Idaho, California, Washington, Colorado, and New York, with some ruling for the plaintiffs and some ruling for the defendants. By 1982, school finance reform activity had tapered off.

A resurgence in school finance litigation began with four cases in 1989 and 1990. In Montana, the court decided for the plaintiff, holding that the state had to reduce financial disparities to guarantee equality of educational opportunity; a state constitutional requirement (*Helena Elementary School District v. State*, 769 P. 2d 684 [Mont. 1989]). In Kentucky, the court also decided for the plaintiff, holding that the constitutional requirement of an "efficient" educational system suggested the need for the state legislature to do more to raise student achievement in poor school districts, particularly by increasing resources to those districts (*Rose v. Council for Better Education*, 790 S.W. 2d 186 [Ky. 1989]). In Texas, the court decided for the plaintiff; this decision also was based on an "efficiency" clause in the constitution (*Edgewood Independent School District v. Kirby*, 777 S.W. 2d 391 [Texas 1989]). New Jersey's Supreme Court held that the constitutional provision of a "thorough and efficient" education in the state necessitated a further reduction in disparities beyond the level the legislature had provided for in its legislation responding to the lawsuits of the 1970s (*Abbott v. Burke*, 575 A. 2d 359 [N.J. 1990]). Currently, some type of school finance litigation is active or in effect in all but 10 states. In Delaware, Hawaii, Iowa, Mississippi, Nevada, Utah, and Vermont, there are no court cases; in Indiana, the case was withdrawn; and in Oklahoma and Kansas, cases are currently dormant.

In addition to engaging in litigation, states have been active in reducing disparities through legislation. Many of these activities have been in response to litigation. In New Jersey, for instance, the legislature passed the Quality Education Act to comply with the 1990 State Supreme Court decision. The court later found the act inadequate, and another act was passed in 1994 (The New York Times 1994). In Texas, legislation passed in response to the lawsuit was found to be inadequate by the courts, and the legislature was called on to pass better legislation (Celis 1994a). In other states, legislation was passed in response to crises in school systems, independent of any litigation. Michigan passed

the best-known legislation of this type in 1994. A school district had been forced to close because of lack of funds. The governor proposed that the state property tax be abolished and replaced with a state sales tax increase and that the additional revenue from this change should be used to increase funds for poor school districts. The proposal also called for capping the level of spending in affluent school districts. The proposal was passed with bipartisan support (Celis 1994b).

Most of this litigation and legislation made no distinctions about the ability of different types of spending to affect student achievement, assuming that increases in total spending by school districts would improve the quality of schools. In the Michigan legislation, districts are guaranteed a minimum level of money for each pupil enrolled in their schools. This guarantee has resulted in higher levels of revenue for many poor school districts. No requirements were made regarding how that money was to be spent, however. In New Jersey, the most recent State Supreme Court decision emphasized the importance of using state equalization funds to improve the level of student achievement, but it did not require the legislature to earmark funds for expenses most likely to accomplish this goal. (Texas also took this approach for the most part.) Educational researchers have found that these court cases have resulted in increases in aggregate spending, with few changes in the spending priorities of school districts (Firestone et al. 1994; Picus 1994).

Kentucky has been the one exception. In *Rose v. Council for Better Education*, the State Supreme Court linked school finance reform to curriculum and governance reform. It called for the establishment of local governance councils for school districts, greater delegation of authority from school districts to schools, creation of state curricular standards, and statewide assessments to measure student progress toward these standards. The court also required the state legislature to earmark dollars for specific activities the court deemed important for raising student achievement, such as instructional materials and professional development programs for teachers. *Rose* even offered financial incentives to districts that could produce large gains in student achievement. Unlike the other cases, *Rose* did distinguish among different types of spending, earmarking funds for those types most likely to raise achievement (Adams 1994; Odden 1993).

While past legislation and litigation usually has conformed to the traditional approach, there may be a move toward a greater emphasis on productivity. The earlier cases relied on aggregate expenditures and have not had great success in increasing achievement. The more recent cases in Texas and New Jersey did not formally earmark spending to raise student achievement; however, they did begin to acknowledge that raising achievement was the goal and that the success of the current finance formulas would be evaluated on the basis of their ability to raise achievement in the poor school districts. As a result, school systems such as Newark's are being forced to go beyond spending their increased

resources in conventional ways. They are finding that, to raise achievement, they need to cut waste, fraud, and abuse and to invest in high-productivity areas. Kentucky went furthest in actually ordering the allocation of resources to high-productivity areas and in conforming to the productivity approach of decentralizing authority to the school level.

Before abandoning the traditional approach completely, however, it may be worthwhile to determine whether some of the more conventional inputs (which the traditional approach increases) are also productive investments. Prior research on these traditional inputs, however, does not provide much guidance on this question.

PRIOR RESEARCH ON THE SPENDING-ACHIEVEMENT RELATIONSHIP

Since the publication of the Coleman Report, educational researchers have undertaken many studies to measure the impact of economic inputs on academic achievement. The studies generally have employed a methodology known as "production function" research, which seeks to measure the amount of each school resource that will maximize student achievement levels and other educational outcomes.

It usually applies a statistical technique known as regression analysis to data bases of students or schools and measures the relationship between various economic inputs and academic achievement, while taking into account background characteristics of students and organizational characteristics of schools.* The reason for taking these additional characteristics into account is that they may explain part of the difference in achievement between high-spending and low-spending school districts. The Coleman Report, for instance, found that the relative affluence or poverty of students' families accounts for differences in achievement; differences in the level of school resources did not have much of an effect beyond this. The inputs such studies measured have ranged from pure spending measures (such as per pupil expenditures) to the types of services these expenditures buy (such as teacher-student ratios and teacher salaries). The results of these studies have been mixed; they have fueled, rather than resolved, the debate on whether or not money matters to educational achievement.

Meta-analyses have attempted to summarize the findings of these studies. Meta-analyses by Hanushek (1989) and Hedges et al. (1994) identified 38 studies, conducted between 1967 and 1987, that examine the relationship between economic resources and student achievement. These studies contained a total of 187 estimates of relationships between resources and achievement when taking into account student background. These estimates (generally contained in equations) measure the impact of seven inputs: (1) per pupil expenditures (55 equations, 11 studies); (2) teacher experience (131 equations, 25 studies); (3) teacher education (88 equations, 18 studies); (4) teacher salary (43 equations, 10 studies); (5) teacher-student ratios (116 equations, 23 studies); (6) administrative inputs (35 equations, 6 studies); and (7) facilities (77 equations, 17 studies).

The meta-analyses diverged regarding the meaning of these studies. Hanushek noted that, for each input, from 7 percent to 29 percent of the relationships to educational outcomes were positive and statistically significant, while for the expenditure measure (per pupil expenditures), 20 percent of the relationships showed a positive significant relationship. Hanushek concluded that "there is no strong or systematic relationship between school

* For reviews of the issues in production function research, see Monk (1992) and Fortune and O'Neil (1994).

expenditures and student performance (Hanushek 1989: 47)."

In another meta-analysis of the same studies, however, Hedges et al. (1994) found that an argument can be made for a positive relationship between school expenditures and educational performance. They noted that, while only a minority of relationships indicate a positive, statistically significant relationship, more do so than would be expected from a random sample of studies. If no association existed between spending variations and achievement variations, it would be expected that only five percent of the relationships would be significant and that this five percent would include both positive and negative significant relationships. Yet, even the inputs with the lowest percentage of positive significant relationships show more than five percent of them to be significant. Furthermore, if no association existed between spending variations and achievement variations, it would be expected that the statistically insignificant relationships would be divided fairly evenly between positive and negative ones. Yet most of the insignificant relationships are positive across inputs, with as much as 70 percent of the relationships between per pupil expenditures and student performance being positive. From this and other evidence, Hedges et al. concluded that the meta-analyzed studies indicate that school spending and achievement are associated with one another (Hedges et al. 1994:13). Hanushek then responded to the arguments of Hedges et. al. (1994), noting that while their meta-analysis provided evidence that relationships between spending and achievement sometimes exist, it still did not constitute evidence of a "strong" or "systematic" relationship, because so many of the studies evinced either no relationship or a negative relationship between the two (Hanushek 1994).

Meta-analyses of other samples of production function studies have been similarly inconclusive. Glass and Smith (1979) meta-analyzed 725 relationships from 77 studies of teacher-student ratios. They concluded that the studies indicate that teacher-student ratios are associated with academic achievement, finding a positive relationship between teacher-student ratios and student achievement for nearly 60 percent of the studied relationships. Odden (1990:224), however, noted that the magnitude of these effects is so small that it is not feasible to increase teacher-student ratios enough to have a significant impact on student achievement.[7]

To some degree, this lack of consensus among the meta-analyses reflects the methodologies of the original studies. First, most of the studies, unlike the Coleman Report, were not nationally representative. Most studied an individual state or school district. This hampers development of a consensus, because different regions of the country

[7] In the area of class size, there have been a few controlled experiments. In these experiments, students were randomly placed in large and small classes and their achievement compared. The most comprehensive of these has found a negative relationship between class size and achievement (Finn and Achilles 1990).

may have different spending patterns and different relationships between these spending patterns and student achievement.

Second, the studies did not distinguish among different types of spending. While they measured multiple inputs (such as teacher experience and teacher-student ratios), the only expenditure measure used was aggregate per pupil expenditures. Using such a gross measure risks missing certain dynamics in the relationship between school spending and academic achievement to the extent that increases in some types of spending have an effect and increases in others do not. For instance, perhaps increased spending on administration does not significantly raise achievement, while increased spending on instruction does. If these types of spending are not measured separately, the apparent effects of spending on instruction will be reduced or eliminated when combined with the lack of effects from administration.

Third, the studies did not take into account how other influences on the process of schooling may mediate between spending and achievement. The effective schools research suggests that certain aspects of the school environment, particularly supportive relations between teachers and principals, positively influence achievement.[8] Yet none of the prior research sought to measure the influence of school spending patterns on school environment.

Fourth, the studies did not all provide rich measures of student background.[9] While the research on measures of the socio-economic characteristics of students indicates that a single measure, known as "socio-economic status (SES)" can be generated by adding together responses to a relatively small number of questions, many of the earlier studies did not include such questions. If SES is poorly measured, it is hard to know if relationships between spending and achievement are, to some degree, attributable to SES differences between students in high- and low-spending districts.

Fifth, most studies did not control for variations in cost between regions. The cost of living in New York City is higher than the cost of living in Montgomery, Alabama. Presumably, this difference means that teachers paid the same dollar amount in the two places are not able to maintain the same standard of living; a dollar will buy less in New York City. As a result, New York City would have to offer higher salaries to successfully recruit the same teachers as Montgomery.[10] Other factors also may

[8] Despite some early criticism of effective schools research (e.g., Cuban 1984; Purkey and Smith 1983), later large-scale multivariate studies have persuaded most researchers that there is a social dimension to school life that plays some independent role in student achievement. The extent of this role, however, is still being debated (Lee, Bryk and Smith 1993).

[9] This was pointed out by Hedges et al. (1994: 12).

[10] When cost of living is taken into account, differentials in per pupil expenditures between high-spending and low-spending states decrease markedly, indicating that states with fewer resources often tend to be states with lower costs of living (Barton et al. 1991).

BEST COPY AVAILABLE

influence the cost of hiring comparable teachers, including the existence or lack of union pressure to increase wages and the overall quality of life in the region. Although most studies did not take these factors into account, they may be as important to include as SES, because differences in achievement between two districts may, to some degree, be due to differences in how much it costs to hire teachers.[11]

Finally, many of the measures of achievement that earlier studies used were not very sophisticated. Some did not use achievement measures at all, but used proxies (such as graduation rates). Some used measures as simple as whether a student passed a minimum competency test. Most did not use measures that took into account the existence of floor and ceiling effects; this has often been found to be an important issue in the measurement of achievement. When students tend to obtain high scores on examinations, they may in fact receive lower scores than would be thought based upon what they know. This is known as a ceiling effect. The floor effect is a similar phenomenon, but in reverse; it occurs for students who tend to obtain low scores on examinations. Only in the past decade, through techniques such as Item Response Theory (IRT), has the ability of ceiling and floor effects to contaminate results been reduced.[12]

Because of the shortcomings mentioned above, prior research into school spending and academic achievement has produced an unclear picture of the relationship between the two. Because the studies did not specify measures of school environment, the effect of school spending on achievement as mediated by environment remains unstudied. Because the studies did not distinguish among different types of spending, the effects of some types of spending differences may be canceled out by the lack of effects of other types. Because some of the studies do not adequately measure socio-economic status, or do not measure the cost of teachers at all, some spending effects may be attributable to these factors. The lack of sophisticated achievement measures may also have affected some results. Finally, the lack of studies drawn from nationally representative data makes it difficult to draw conclusions about spending effects generalizable to the United States as a whole. This last problem is in large part due to the lack of a nationally representative data base of school spending patterns and student achievement, since data on school spending and achievement typically have been collected in separate data bases, for different samples of students and schools. To address the problem of national representativeness, this study has had to construct a nationally representative data base from various data sources (this is discussed in the following section).

[11] Fortune and O'Neil (1994: 24) pointed this out.

[12] For a discussion of this shortcoming in production function research, see Fortune and O'Neil (1994: 24). For a discussion of IRT, see Hambleton et al. (1991).

DATA SOURCES AND STUDY METHOD[13]

To relate education spending to academic achievement and to address the other methodological issues presented earlier, the study needed to collect data on student academic achievement, types of per pupil expenditure for the school district, the social environment of the school, teacher-student ratios and teacher education levels in schools, the socio-economic status of students, and the cost of education in the region.

Since no data base contains all of this information, data were be drawn from three sources: (1) the National Assessment of Educational Progress (NAEP); (2) the Common Core of Data (CCD); and (3) the Teacher's Cost Index (TCI) of the National Center for Education Statistics (NCES). A description of each data base and the information collected from it for this study follows.

NAEP is administered every two years to nationally representative samples of fourth, eighth, and twelfth graders. It consists of tests in a variety of subjects, not all of which are included in each administration. For instance, 1992 included tests in mathematics and reading but not in geography (which was included in the 1994 administration). For this study, the 1992 mathematics assessment was selected. NAEP also includes background surveys administered to students, principals, and, for the fourth- and eight-grade surveys, to teachers. The information from the background surveys makes it possible to measure the social environment, teacher-student ratios, and the average teacher education level in the school, as well as the average socio-economic status of its students. Because the twelfth-grade survey does not have a teacher component, it is not possible to measure social environment for this grade level. Consequently, only fourth and eighth graders were analyzed in this study.[14]

The Common Core of Data is administered by the National Center for Education Statistics (NCES) every year to all public school districts in the nation. The districts report to NCES the amount spent in a variety of categories, including capital outlays, instruction, district-level administration, school-level administration, and support services (such as guidance counseling and transportation to and from school). The districts also report their enrollments, which makes it possible to calculate per pupil spending in each area. The financial information from the 16,666 school districts in academic year 1991-1992 was chosen for analysis because it represents the spending in the academic year that would have had the most immediate effect on mathematics achievement in 1992.

The Teacher's Cost Index was developed by NCES from a study of the staffing of schools and the cost characteristics of the district in which they are located. It takes into account differences in the cost of living

[13] For full discussion of technical details of this study, see the Appendix and Wenglinsky (1996b).

[14] For an overview of NAEP, see Johnson (1994).

between different regions, as well as variations in the cost of teachers due to union arrangements, labor laws, employment trends in comparable occupations, and the quality of life. Therefore, it is known as a "cost-of-education" (rather than a "cost-of-living") index. For this study, the index at the state level was used, as district comparisons of the cost of education were not readily available.[15]

The study linked together the three sources of information for fourth graders and eighth graders to form one data base for each of the grades. Because the Common Core of Data (CCD) information is collected at the school district level, analysis had to be conducted at that level. The student and school characteristics in the National Assessment of Educational Progress (NAEP) were averaged out to their district-level averages separately for each grade level. For each NAEP district, the district-level identification number was matched to an identical identification number in CCD. Most of the NAEP districts for which there was no match were private schools, which are not required to provide financial information for CCD. Of the rest, eight fourth-grade districts and five eighth-grade districts were matched through comparing their names and addresses between CCD and NAEP. For all districts, the appropriate cost index score for the state in which the school district is located was added to the data bases. The resulting data bases consisted of 203 fourth-grade districts and 182 eighth-grade districts.

For each grade, data were analyzed in three steps. First, a basic description of the students in this nationally representative sample was calculated from the national averages of the characteristics of interest (for example, finance information, achievement information, social environment). This information is important both to illustrate the financial characteristics of the nation's school districts and to test whether or not the collected information is comparable to that found from other sources.

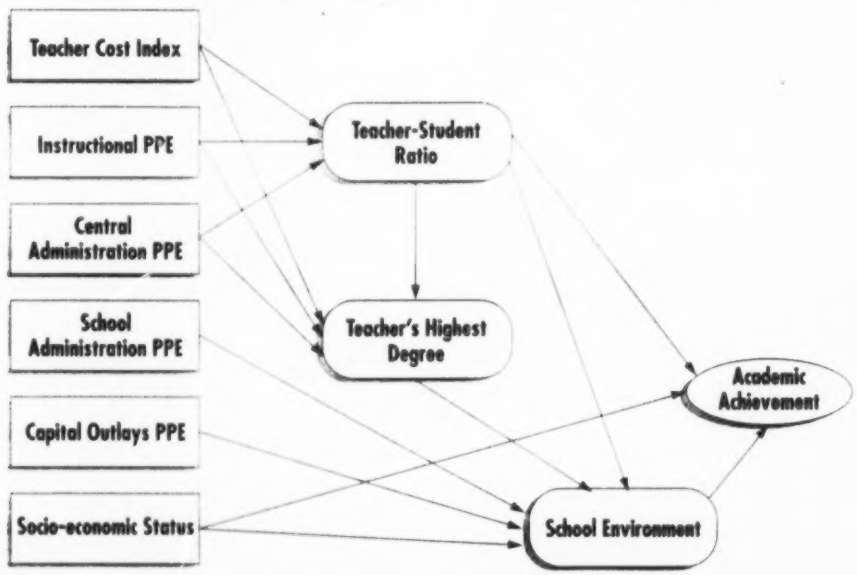
Second, a software program known as LISREL 8 was used to test whether or not there were statistically significant relationships between the characteristics described. It looks for such relationships not only between financial characteristics and achievement, but also between all of the characteristics, because financial characteristics might affect achievement only because of an intervening characteristic (Hayduk 1987). For instance, perhaps money spent on instruction improves the school environment, and this improvement in school environment in turn raises achievement. LISREL provides a flow chart of how various resource allocations filter down into achievement gains. In testing relationships, it also takes into account all of the other relationships. For instance, if a supposed relationship between spending and achievement is attributable to differences in

[15] For a discussion of the theory of cost-of-education indexes, see Barro (1994). For the development of the TCI, see the National Center for Education Statistics (1995).

socio-economic status (SES) between school districts, it would treat the spending-achievement relationship as not significant but treat that SES-achievement relationship as significant. Figure 1 presents all of the possible flows of dollars that were tested by this study for fourth and eighth graders using LISREL. A caveat to the use of LISREL to test relationships between spending and achievement should be noted. LISREL analyzes variations in the characteristics measured by the data base. It therefore provides information on the degree to which, within the range of current practice, changes in one characteristic are associated with changes in another. For instance, the finding of a relationship between socio-economic status and achievement indicates that, within the existing limits of affluence, increases in socio-economic status are associated with increases in achievement. This caveat to LISREL excludes two possible interpretations of the existence or lack of a relationship. First, relationships do not indicate that one particular absolute level of a characteristic is associated with a particular achievement score. Second, relationships only apply within the bounds of the highest and lowest values of each characteristic. If spending were to be unassociated with achievement, it would not imply that zero dollars of spending will have no effect on achievement. Rather, it suggests that incremental decreases do not have an effect.

LISREL also provides information on the power of the relationship between these characteristics (e.g., how many dollars translate into how many more teachers per student), but this information is of limited use because it makes many assumptions about the data which may not be true.

FIGURE 1: HYPOTHESIZED PATHS TO ACHIEVEMENT



BEST COPY AVAILABLE

To quantify the strength of the effect of characteristics on each another, the school districts were divided into subgroups and comparisons made between them. First, districts were divided into four groups: (1) districts of below-average socio-economic status (SES) students and below-average teacher costs; (2) districts of below-average SES students and above-average teacher costs; (3) districts of above-average SES students and below-average teacher costs, and (4) districts of above-average SES students and above-average teacher costs. This was done because some differences in achievement may be attributable to SES and teacher cost; therefore, only districts that were relatively homogenous in these respects were compared. Second, each of these four types of district was compared with another for each relationship LISREL found significant. (For instance, if instructional spending and teacher-student ratios are significant, the teacher-student ratios are compared between school districts with above-average instructional spending and those with below-average instructional spending. This provides a sense of how much of a difference instructional spending makes to teacher-student ratio).[16]

The data and procedures outlined here were designed to address the drawbacks of prior research. The analysis deals with different types of spending, rather than simply aggregating spending (as had been done in most previous work). It includes measures of social environment and is set up to test the notion that social environment intervenes between spending and achievement. The analysis includes sophisticated measures of achievement that take into account, to some degree, floor and ceiling effects. It includes strong measures of socio-economic status and the regional cost of education and (through LISREL), takes these measures into account in testing relationships between spending and achievement. Finally, the data represent the most up-to-date, nationally representative information on both spending and achievement.

[16] The comparison of subgroups is not as statistically rigorous as estimates using LISREL or other multivariate techniques, but it is useful for illustrative purposes. It also does not make the statistical assumptions that LISREL and most other techniques require. For another example of the use of subgroup comparisons to address deficiencies in production function models, see Fortune and O'Neil (1994).

FINDINGS

Description of School Districts

Looking first at the descriptive spending characteristics for fourth graders (Table 1), it appears that the information conforms broadly to what is known about school spending.

The national average per pupil expenditures on instruction are $2,973.89 a year, or about 60 percent of total district spending on education. Instruction includes the salaries and benefits for teachers and their aides, as well as instructional materials. Other research has found that these amounts vary little; most school districts spend from 55 percent to 65 percent on instruction. Expenditures on school district (central office) administration are $110.72 per pupil, or less than three percent of total district spending on education. At first, given the discussion of the overwhelming cost of administration as portrayed in the media, this amount may appear surprising. The notion of high administrative costs, however, conflates two very different kinds of expenses—those for support services for students and schools (transportation, food, janitorial services, counseling) and those to pay the administrators who formulate and administer district policy (the superintendent and his or her staff). The latter generally has been found to be around five percent of district spending and is what is measured here as central office administration. Support service spending generally has made up 30 percent of total expenditures, and it does so in this data base of fourth graders. Expenditures on school administration (the principal and his or her staff) are $283.70, about six percent of total expenditures. Contrary to the notion that the administration of schools is "top-heavy" more is spent per pupil on administration at the school level than at the district level. Finally, expenditures on capital outlays (building and repairing facilities) is $486.89. This amount usually is not categorized as a part of total district spending, but it conforms in amount to what research has found.

The expenditure patterns of school districts of eighth graders are not markedly different.[17] On average, slightly more is spent per pupil in the eighth grade sample than in the fourth grade sample. The one exception is central office administration, where slightly less is spent in the eighth grade sample. Thus, the data for both grades conform to what is known from other research.[18]

[17] Because expenditure data are on the district level, expenditure differences between the fourth grade and eighth grade samples are differences in total spending on all grades; therefore, they cannot be used as comparisons of spending between fourth and eighth graders. All other comparisons can be understood as comparisons between fourth and eighth graders because the data are on the school or student level.

[18] For examples of distributions of resources found in other studies that conform to those found here, see Adams (1994) and Miles (1995). For an example of a press report that does not distinguish between the different types of administrative expenditure, see Barnabel (1994).

TABLE 1: DESCRIPTIVE CHARACTERISTICS OF FOURTH AND EIGHTH GRADERS

Characteristic	Fourth Graders (# School Districts =203)	Eighth Graders (# School Districts=182)
Teacher's Cost Index	98.6	98.27
Per Pupil Expenditures (PPE) (in Dollars)		
Instructional PPE	$2,973.89	$3,043.66
Central Office Administration PPE	110.72	104.60
School Administration PPE	283.70	289.14
Capital Outlays PPE	486.89	487.90
Socio-economic Status Scale		
Encyclopedia in Home	.71	.77
Family Gets Magazine	.72	.80
Family Gets Newspaper	.73	.76
More than 25 Books in Home	.92	.95
Mother's Education (1= < H.S.; 2=H.S. Graduation; 3=Some College)	2.06	1.68
Father's Education(1= < H.S.; 2=H.S. Graduation; 3=Some College)	2.14	1.79
Students Receiving School Lunch (4=11-25%; 5=25-60%)	4.69	4.34
School Resources		
Teacher-Student Ratio	.05	.06
Teacher's Highest Degree (2=Bachelors, 3=Masters)	2.54	2.56
School Social Environment Scale		
Student Tardiness	3.16	2.92
Student Absenteeism	3.15	2.92
Teacher Control over Instruction	2.97	2.77
Teacher Control over Course Content	3.14	2.65
Regard for School Property	3.43	1.80
Teacher Absenteeism	3.58	3.34
Student Class Cutting	3.97	3.51
Mathematics Achievement	210.65	262.72

While no large differences in spending patterns exist between the two samples, there do appear to be large differences in the socio-economic characteristics of students in the two grades. Socio-economic status (SES) is measured by examining a series of characteristics of students' families: resources available in the home (books and periodicals), the education levels of the parents, and the percentage of the student body in the student's school poor enough to receive a reduced-price or free school lunch. For both grades, from 70 percent to 80 percent of all students live in homes with

encyclopedias, magazines, and newspapers. A higher percentage of students (more than 90 percent), possess more than 25 books in the home. Some difference exists between the grades, however. Eighth graders are more likely than fourth graders to have these resources. Eighth graders are also in schools where fewer students are poor enough to qualify for a school lunch. For the education measures of SES, the proportions are reversed. The parents of fourth graders have higher levels of educational attainment than those of eighth graders; most parents of fourth graders have completed high school, while most parents of eighth graders have not.

Regarding school resources (as opposed to the actual dollars spent), differences exist in teacher-student ratios but not in teacher education. Classes tend to be somewhat smaller for eighth graders than for fourth graders; the average teacher-student ratio is .05 teachers per student (or one teacher for 20 students) for fourth graders, compared with .06 teachers per student (or one teacher for 17 students) for eighth graders. While these teacher-student ratios may appear quite small, they mask larger teacher-student ratios for most students; this is because the teacher-student ratio also includes special education classes, which often have from 1 to 10 students per teacher (Miles 1995). In terms of a teacher's education, there is little difference between fourth- and eighth-grade teachers.

The school social environment of fourth graders seems much more positive than that of eighth graders. Social environment is measured by indicators of student involvement (tardiness, absenteeism, class cutting, regard for school property) and teacher involvement (absenteeism, control over instruction, control over course content). Problems with student and teacher commitment appear to be much less pronounced in the schools of fourth graders than in those of eighth graders. The largest difference is in regard for school property (with much less vandalism in fourth grade), followed by substantial differences in every other category.

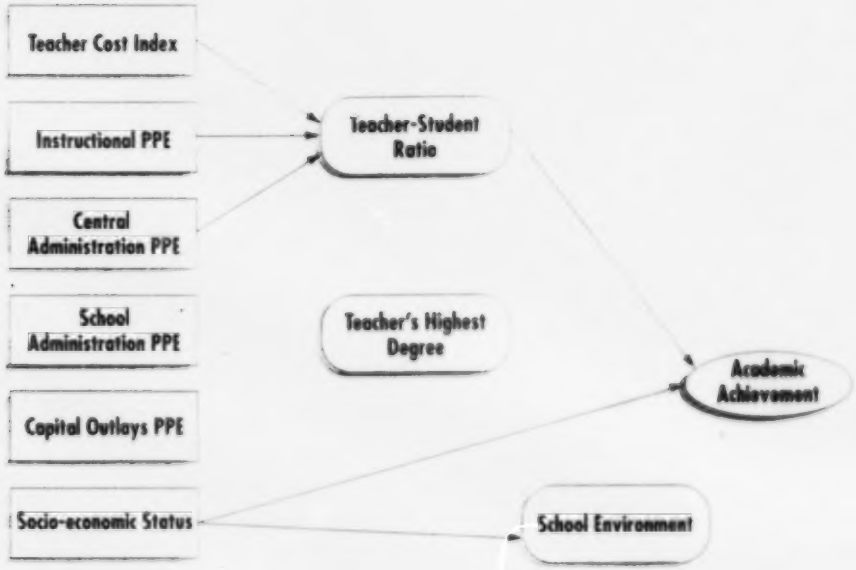
The two other characteristics in the study (the Teacher's Cost Index (TCI) and mathematics achievement) require further explanation. The TCI differs little between the grades, because both are nationally representative samples. Both scores are close to 100, which was the national average as calculated in the TCI study. Mathematics achievement is on a single proficiency scale; this accounts for the lower score for fourth graders (210.65), compared with that for eighth graders (262.72). The single scale permits comparisons not only between fourth and eighth graders, but also between subgroups of them. This makes it possible to answer questions such as: "Do high-socio-economic status (SES) fourth graders with large teacher-student ratios have proficiency scores that are much lower than those of low-SES eighth graders with small teacher-student ratios?"

BEST COPY AVAILABLE

Testing Relationships

The findings for each grade confirm the notion, advanced earlier, that variations in some types of conventional spending have an effect on achievement, while variations in others do not. For fourth grade (Figure 2), the flow of dollars to achievement can be seen moving left to right. Increases in school administration expenditures and capital outlays do not appear to raise achievement or any of the intervening characteristics that might themselves raise achievement (teacher-student ratios, teacher's highest degree, school environment). On the other hand, increases in instructional spending and central office administration spending both appear to raise teacher-student ratios. In other words, more spending on instruction does allow school districts to hire more teachers for each student, and more spending on central office administration makes it more likely that they will spend money in this manner. Increases in teacher-student ratios, in turn, appear to influence academic achievement positively. Variations in the teacher's highest degree do not appear to be associated with any spending measures or with the achievement of fourth graders. School social environment also does not appear to play a role. The Teacher's Cost Index (TCI) and socio-economic status (SES) are important influences in this process, independent of spending.

FIGURE 2: PATHS TO FOURTH-GRADE ACHIEVEMENT

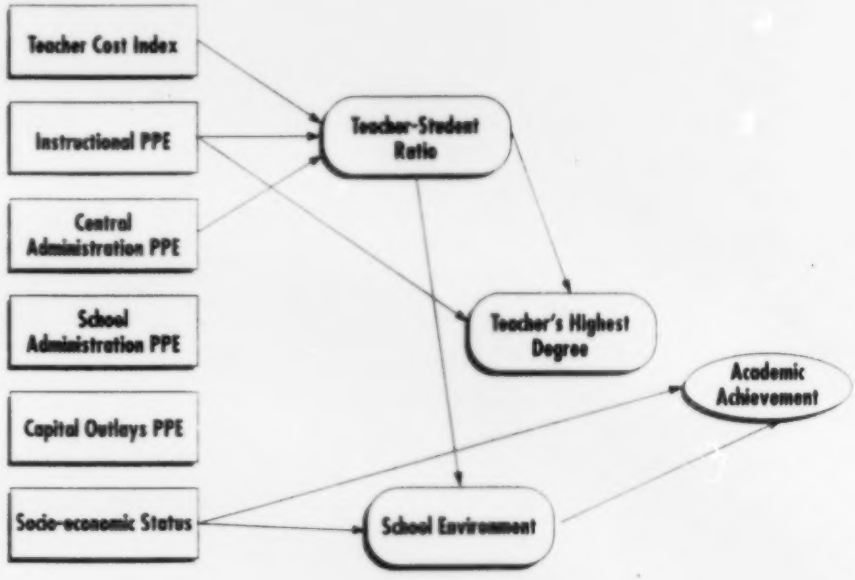


BEST COPY AVAILABLE

For eighth graders, the flow of dollars is more complex (Figure 3). Again, variations in spending on school administration and capital outlays do not play a role. Increases in spending on central office administration and instructional spending again influence teacher-student ratios. Increased teacher-student ratios, however, do not directly result in achievement gains. Rather, high teacher-student ratios improve the school environment, which in turn raises academic achievement. Another difference is that, while the teacher's highest degree still does not influence academic achievement, it is influenced by instructional spending. The greater the level of per pupil expenditure on instruction, the higher the teacher's education level. As in the past, socio-economic status (SES) and the Teacher's Cost Index (TCI) prove to be important factors.

Thus, there appear to be different paths from spending to achievement for fourth and eighth graders that have an effect above and beyond the SES of students and variations in the cost of education between states. For fourth graders, increased spending on instruction and central office administration raises teacher-student ratios, which raises mathematics achievement. For eighth-graders, increased spending on instruction and central office administration raises teacher-student ratios, which improves the school social environment, which in turn raises mathematics achievement. The quantification of these effects, however, requires a comparison of subgroups for each of these steps in the flow of resources.

FIGURE 3: PATHS TO EIGHTH-GRADE ACHIEVEMENT



BEST COPY AVAILABLE

Comparing Subgroups

Table 2 provides for fourth graders the estimation of how much each step in the process affects another step. Above-average instructional expenditures appear to increase the number of teachers per student from .056 (17.9 students/teacher) to .0564 (17.8 students/teacher) for low-socio-economic status (SES), low-cost regions; from .0421 (23.8 students/teacher) to .0526 (19.0 students/teacher) for low-SES, high-cost regions; from .0518 (19.3 students/teacher) to .0642 (15.6 students/teacher) for high-SES, low-cost regions; and .0445 (22.5 students/teacher) to .0523 (19.1 students/teacher) for high-SES, high-cost regions. On average, the increase is .008 teacher per student, with the highest increases for low-SES, high-cost regions and high-SES, low-cost regions. For central office administration expenditures, the differences conform to the same pattern, with increases in all four subgroups but with the highest increases in the same two groups as for instructional expenditures. The effect is slightly smaller for central office administration, however, with an average difference of .006, compared with the .008 difference for instruction.

Differences in teacher-student ratios translate into substantial differences in achievement for the four groups. The achievement differences are largest for high-SES, high-cost regions and low-SES, high-cost regions, with increases of more than eight points. High-SES, low-cost regions had the least change (less than two

TABLE 2: COMPARISONS OF SUBGROUPS: FOURTH GRADERS

	Teacher-Student Ratios				Achievement Scores	
	Low Instructional Expenditures	High Instructional Expenditures	Low Central Office Expenditures	High Central Office Expenditures	Low Teacher-Student Ratio	High Teacher-Student Ratio
Low-SES, Low-cost Regions	.0560	.0564	.0554	.0566	196.74	200.79
Low-SES, High-cost Regions	.0421	.0526	.0467	.0536	195.52	204.53
High-SES, Low-cost Regions	.0518	.0642	.0512	.0615	218.80	220.34
High-SES, High-cost Regions	.0445	.0523	.0480	.0530	219.08	227.73

BEST COPY AVAILABLE

points). These point differences are between below average and above average teacher-student ratios, with a six-point difference representing approximately half a grade level. In other words, fourth graders in smaller-than-average classes are about a half a year ahead of fourth graders in larger-than-average classes.

For eighth graders (Table 3), increases in teacher-student ratios are more marked. Above-average instructional expenditures translate into increases from .0561 teachers per student (17.8 students/teacher) to .0804 (12.4 students/teacher) for low-SES, low-cost regions; from .0536 (18.7 students/teacher) to .0632 (15.8 students/teacher) for low-SES, high-cost regions; from .0550 (18.2 students/teacher) to .1091 (9.2 students/teacher) for high-SES, low-cost regions; and from .0530 (18.9 students/teacher) to .0669 (14.9 students/teacher) for high-SES, high-cost regions. Central office expenditures conform to a similar pattern, with an average difference of .0178 between high- and low-spending districts, with the highest difference for high-SES, low-cost regions and the lowest difference for low-SES, high-cost regions. Teacher-student ratios, in turn, appear to produce marked differences in social environment, averaging a .477 point difference in the social environment scale, with the greatest difference in low-SES, low-cost regions, the next greatest difference in low-SES

TABLE 3: COMPARISON OF SUBGROUPS: EIGHTH GRADERS

	Teacher-Student Ratios				Social Environment		Achievement Scores	
	Low Instructional Expenditures	High Instructional Expenditures	Low Central Office Expenditures	High Central Office Expenditures	Low Teacher-Student Ratio	High High Teacher-Student Ratio	Low Social Environment	High Social Environment
Low-SES, Low-cost Regions	.0561	.0804	.0519	.0622	3.639	4.568	247.34	252.70
Low-SES, High-cost Regions	.0536	.0632	.0588	.0656	4.718	5.354	238.99	259.72
High-SES, Low-cost Regions	.0550	.1091	.0548	.0974	6.077	6.336	268.46	273.70
High-SES, High-cost Regions	.0530	.0669	.0573	.0688	6.371	6.454	273.71	274.07

BEST COPY AVAILABLE

high-cost regions, the next greatest difference in high-SES, low-cost regions, and the smallest difference in high-SES, high-cost regions.

Differences in social environment translate into large differences in achievement, comparable to the differences attributable to teacher-student ratios among fourth graders. The largest difference, again, is for low-cost, high-SES regions (more than 20 points), while the smallest difference (an insubstantial one) is for high-SES, high-cost regions. The other two groups are in the middle, with differences based on social environment of slightly more than five points.

In sum, it seems that there is a substantial flow of dollars to achievement. Dollars spent on instruction and central office administration do not disappear but, in fact, substantially raise teacher-student ratios. These larger teacher-student ratios, in turn, result in achievement gains in mathematics (although these gains vary, depending on the socio-economic status of students and the costliness of the region of the country in which they live). The largest effects seem to be for poor students in high-cost areas.

BEST COPY AVAILABLE

Interpretations

The difference in the flow of dollars between fourth and eighth graders is interesting. For fourth graders, it appears that increased teacher-student ratios result directly in achievement gains; for eighth graders, increased teacher-student ratios raise achievement only through improving the school social environment. The different level of positive social environment for each grade offers a possible explanation for this difference. As Table 1 indicated, fourth graders are less likely than eighth graders to show problems in their social environments. Fourth graders are less likely to damage school property or to be late, absent, or cut class, and their teachers are less likely to be absent. This suggests that a negative social environment is not much of an obstacle to fourth graders, and, therefore, the benefits of small classes are entirely pedagogical. In eighth grade, however, there is a higher incidence of negative social environmental behaviors, and this creates a barrier to high achievement among students. This can be seen most clearly in the large differences in achievement between negative and positive social environment school districts in Table 3 (differences are more than half a grade level). Thus, the benefit of small classes for eighth graders is their ability to reduce the tendency of students and teachers to show these problem behaviors. Fewer students mean that teachers can be more attentive to potential social problems, and teachers themselves may feel more attached to their students. For both fourth and eighth graders, then, the benefit of an increased teacher-student ratio is its ability to promote social cohesion. For fourth graders, this social cohesion is primarily pedagogical; for eighth graders, it is primarily geared toward preventing negative social behaviors.[19]

The relationship of central office administration to education spending is a second interesting matter of interpretation. The fact that increased instructional spending increases teacher-student ratios should not be a surprise. It indicates that additional resources devoted to instruction often are spent on increasing teacher-student ratios, a conventional budgetary decision. Both models also indicate, however, that when more money is spent on administration at the district level, teacher-student ratios are increased. In contrast, additional dollars spent on administration at the school level are not associated with increases in teacher-student ratios or, indeed, with any other educational characteristics in the model. This suggests that larger, better-financed district offices may be more successful than principals or other administrators

[19] Another possible interpretation of the mediating effect of social environment is that it is an interaction effect. If this is the case, then, for eighth graders, the significant relationships among teacher-student ratio, social environment, and achievement suggest that higher teacher-student ratios only affect achievement if they are accompanied by improvements in the social environment of the school. Recent small-scale research (Murnane and Levy 1995) found this to be true for 15 schools in Austin, Texas.

BEST COPY AVAILABLE

in ensuring that dollars actually reach the classroom.[20]

Finally, another interesting finding is that, while variations in teacher-student ratios are associated with variations in achievement, the variations in teacher education are not. This suggests that not all spending on instruction is of equal worth in promoting high achievement. Among eighth graders, additional dollars spent on instruction take two paths—to increases in teacher-student ratios and to increases in teacher education. Of these paths, only that of improving teacher-student ratios ends with increases in achievement. Central office administration is not associated with increases in teacher education, however, suggesting that well-financed district administrations are no more likely than their less well-financed counterparts to invest their dollars in the less effective of these two directions.[21]

---

[20] This interpretation should not be taken to mean that increases in spending on principals' offices cannot have a positive effect on achievement but, rather, that additional dollars will have more of an effect if spent on central office administration. It may be that school-level administration spending has a positive relationship to achievement that is much weaker than that of central office administration but that it is also too weak to be identified in this district level of analysis, due to the statistical issue of "aggregation bias." In addition, in many cases, the budget of the principal's office is decided entirely at the central office level. Some of the lack of association between variations in spending on the principal's office and variations in teacher-student ratios, therefore, may be the responsibility of the central office and not of the principal's office. These possibilities would not change the policy implications of the findings as discussed here, however, because it is still true that the money allocated unambiguously to the central office is strongly associated with teacher-student ratio increases.

[21] While the study shows that variations in teacher education are not associated with variations in achievement, this is not to say that other measures of teacher quality may not prove important. Teacher experience, teacher education in the subject matter being taught, and teacher cognitive ability are not measured here. For a complete discussion of the interpretation of these findings and associated methodological caveats, see Wenglinsky (1996b).

BEST COPY AVAILABLE

Knowing More

As the matters of interpretation just discussed indicate, the LISREL models present at best a schematic picture of the role of finance in the schooling process. Additional steps should be taken to paint a fuller picture. First, research needs to be conducted at the student and school levels. The current study measures differences at the school district level and may mask intradistrict differences between schools and intraschool differences between students. This problem can be addressed in part through examining the relationship between nonexpenditure school resources and student achievement. While per pupil expenditures were drawn from the Common Core of Data (CCD), which collects data at the district level, teacher-student ratios and teacher education were drawn from the National Assessment of Educational Progress (NAEP) and therefore are available for each school. Conducting nonexpenditure resource analyses at the school level, then, could measure the degree to which the teacher-student ratio to achievement relationship applies to different types of school and student.

A second shortcoming of the current study is that it is limited to public school finance. More than 10 percent of all students in the United States attend private schools, and it is possible that the spending-achievement relationship is different for these students. Unfortunately, while NAEP does include private schools, CCD does not. National Center for Education Statistics conducts a private school survey, but the survey does not include much financial information. Thus, while nonexpenditure resource issues can be investigated using NAEP alone, the influence of per pupil expenditures cannot be investigated in this way.

The lack of expenditure data at the school level raises a third issue—the need for a data base that includes this information, as well as information on achievement, student characteristics, and school characteristics. Perhaps the easiest way to produce such a data base would be to add finance questions to the NAEP school background questionnaires. Doing so would permit school-level analyses of the relationship between spending and achievement, as well as public-private comparisons.

Moreover, the current study analyzes cross-sectional data. The use of cross-sectional data to study issues that make assumptions about cause and effect is controversial.[22] To strengthen the findings of the current study, it would be worthwhile to analyze a longitudinal data base that included all of the measures used in this study as well as past student achievement. Because no data base exists that meets all these requirements, and the synthesizing

[22] The use of cross-sectional data in this study may be particularly problematic in the case of capital outlays, since the effects of capital investments on student performance may take years to manifest themselves. Nevertheless, although the use of cross-sectional data to make cause-and-effect statements is controversial, most researchers support its use as a stopgap in the absence of analogous longitudinal data (e.g., Mosteller and Moynihan 1972).

BEST COPY AVAILABLE

of data bases that meet these requirements is highly infeasible, it is unlikely that this problem will be rectified without new data collection efforts.

Finally, it is beyond the scope of this study to suggest the extent to which new earmarked funds should come from the redeployment of existing dollars or new appropriations. This analysis only examines the effects of resource variations within the limits of current practices. If redeployment were to result in cutting funds too deeply for a particular type of expenditure found in this analysis not to be associated with achievement, declines in achievement could result nonetheless.

BEST COPY AVAILABLE

POLICY IMPLICATIONS

Although much remains to be done, some preliminary policy implications can be drawn. The current study provides some support for both the traditional and productivity perspectives on school finance.

It supports the notion of productivity researchers that some investments are not particularly productive for heightened achievement levels among students. Additional expenditures on capital improvements, principals' offices, and teacher education levels do not appear to raise test scores. On the other hand, one investment found to produce heightened achievement levels is fairly conventional—improving teacher-student ratios. In addition, while the productivity approach is extremely concerned with bloated bureaucracies, the current study suggests that a significant fiscal commitment to central office administration is associated with more money being spent on small classes. Thus, while principals may be better positioned than central offices to make instructional and personnel decisions, some budgetary decisions may be better left to the superintendent. Thus, the productivity approach is justified in calling for the careful investment of resources, but some conventional spending methods should not be discarded so quickly.

From the perspective of the courts and state legislatures seeking to reform school finance, these findings support the targeted approach of Kentucky. Equalizing resources without earmarking them for investments most conducive to increased achievement may result in more money being spent but without achieving results. In these fiscally conservative times, and after years of spending increases and aggregate equalizations that have not produced results, courts and legislatures cannot afford to continue to ignore the relative productivity of educational investments. This research also suggests, however (in contrast to the allocations made by the Kentucky Supreme Court), that it would be worthwhile for courts and legislatures to earmark some of these dollars for instructional expenditures and, particularly, for increasing teacher-student ratios. It also suggests that the emphasis of the Kentucky court on increasing the decision-making authority of principals, while perhaps appropriate in instructional and administrative matters, should be applied with great caution in fiscal ones; it may be worthwhile for districts to retain some power in making allocative decisions.

From the perspective of school superintendents, who grapple each day with the problem of scarce resources without being able to count on additional aid from the state or federal governments, this study suggests that greater efforts be made to ensure that dollars get to the classroom, particularly for making classes small. Superintendents should experiment with budgetary strategies to accomplish this with existing resources. One way is to target more dollars for instruction, at the expense of capital projects and support services like custodial

BEST COPY AVAILABLE

work. Another way is to reduce the proportion of instructional dollars being spent on recruiting teachers with high levels of education. With both of these innovations, superintendents would be reallocating some dollars from areas less closely associated with academic achievement to areas more closely associated with academic achievement. These types of innovation may be particularly fruitful for low-SES high-cost school districts, where this study found teacher-student ratios to have a greater effect than for other districts. Through these types of change in budgetary decision making, superintendents may be able to take advantage of the productivities this study uncovered and raise student achievement in the most troubled school districts.

BEST COPY AVAILABLE

APPENDIX: HOW THE STUDY WAS CONDUCTED

Data

The data employed for this analysis are drawn from three sources: (1) The National Assessment of Educational Progress (NAEP); (2) the Common Core of Data (CCD); and (3) a Teacher's Cost Index (TCI).

NAEP is a nationally representative data base of students and schools collected by the Educational Testing Service (ETS) and Westat under contracts from the National Center for Education Statistics (NCES). CCD is a data base consisting of the universe of school districts in the United States, collected by NCES. The TCI was developed by NCES to measure regional variations in the cost of teachers.

NAEP provides information on student achievement, school social environment, student and school socio-economic status, teacher-student ratios, and teacher education levels. It is administered by ETS and Westat every two years to nationally representative samples of fourth, eighth, and twelfth graders and to their teachers and principals. The subject areas tested vary but have included mathematics, reading, history, geography, writing, and science. The information NAEP collects is used to assess what students around the country know, to make comparisons in the levels of knowledge of various regional, ethnic, socio-economic and gender subgroups, and to measure the progress of students in the nation both over time and between grades (see Johnson 1994 for an overview of NAEP; Mullis et. al. 1993 for report card for 1992 Mathematics Assessment).

Certain methodological issues need to be addressed in the secondary analysis of NAEP. NAEP is not administered to a simple random sample of students (where every student has an equal chance of being selected) but to a stratified, clustered sample with differential probabilities of selection. Consequently, data analysis must take into account the characteristics of the sample. For the calculation of parameters from the data base, it is necessary to weight each case by the inverse of the probability of that case being selected from the population. For the calculation of standard errors or other measures of sample variability, it is necessary to account for the weighting, stratification, and clustering of the sample. This is done by using a procedure called "jackknifing," which allows for the appropriate handling of the sample design in the estimation of sampling error. Using weighting and jackknifing allows analyses that are unbiased and that correctly reflect the sample design (Johnson 1989).

A second issue is the measurement of academic achievement. Of all of the questions in the survey designed to measure achievement in a subject area, only a subsample is administered to any given student. Inferences of what a particular student would have scored on questions not administered to him or her, and the resulting total score that the student would have received, are therefore complex. NAEP uses Item Response Theory (IRT) to summarize

BEST COPY AVAILABLE

student performance. Because individual student performance is not well measured, due to the small number of items administered to any student, traditional IRT point estimates of student performance ("scale scores") are inappropriate for estimating population distributions. Special procedures are used to account for the fact that individual proficiencies are not well determined. These procedures, called "plausible values methodology," use the IRT model for the relationship of the item responses to proficiency, a statistical model relating proficiency to background variables, and the observed data (item responses and background variables for each individual) to simulate sets of plausible values for proficiencies. Data analysis proceeds by repeatedly estimating parameters of statistical models, using each individual's plausible values in turn. The final results are based on the average of the individual estimates, with the variability in results for different plausible values providing a measure of the impact of measurement error.

The Common Core of Data (CCD) was used to provide financial information on school districts. Currently there is a dearth of nationally representative financial data on elementary and secondary education. At the school level, the only pure spending measure available for a recent nationally representative sample is total per pupil expenditures, collected in the High School and Beyond study of students and schools in 1980. To analyze different types of spending, then, it is necessary to use data collected at the district level; CCD is ideal for this purpose, because it collects financial information from the universe of public school districts in the country on a yearly basis, distinguishes different types of spending (instructional, central office administration, school-level administration, other), and includes identifying information (such as names and addresses of school districts) that makes it possible to link it with the National Assessment of Educational Progress (NAEP).

The key methodological issue in the use of the Common Core of Data (CCD) is to ensure the validity of the financial data. All data are self-reported by school districts, in response to a National Center for Education Statistics (NCES) questionnaire that asks for spending levels in a variety of budgetary categories. In some cases, the request is unambiguous, and responses conform to what would be expected. In others, the request is ambiguous enough that different respondents will interpret it as seeking different information. CCD can be used to estimate reliably district per pupil expenditures on instruction, central office administration, school-level administration, and capital outlays, because the percentage of spending in each of these areas conforms broadly to what is known to be the national average. In other categories, however (such as spending on transportation and research), responses depend on the budgeting practices of individual schools (particularly their charts of accounts); some enter no expenditure on transportation because they include transportation under another

BEST COPY AVAILABLE

heading. Therefore, analysis of these other spending areas (which can be loosely classified as ancillary services) is not possible using CCD.

Finally, the Teacher's Cost Index (TCI) is the result of a study by NCES. NCES has conducted analyses to develop an index of the cost for a particular region of the country of hiring teachers (NCES 1995). The cost of hiring teachers, even those with similar levels of education and experience, can be expected to vary regionally, because the cost of living, quality of life, and other dynamics of the labor market differ regionally. NCES has developed a cost index, applying regression analysis to the Schools and Staffing Survey, an NCES survey conducted in 1990-1991. The regression analysis estimates the influence of various factors on teacher salaries; these include factors under the control of schools and school districts (such as teacher experience and education), as well as characteristics that are not under the control of schools (such as the regional cost of living and quality of life). The resulting estimates of the impact of these nondiscretionary characteristics on teacher salaries can then be used as estimates of teacher costs in a particular region, holding constant the discretionary factors. NCES has estimated TCI scores for each state, and these are included in this analysis (NCES 1995:51).

To analyze data from these three sources, all needed to be placed at a single level of analysis. The district level was selected because it is the minimum level of aggregation for the CCD. The NAEP data, collected at the individual and school levels, were aggregated to the district level by calculating the mean for all selected variables for each district. Because no questionnaire is administered to the teachers of twelfth graders, only NAEP data for fourth and eighth graders taking the 1992 mathematics assessment were included. This resulted in data for 230 districts. These data were then linked to the 1991-1992 CCD data base, using identification numbers provided by Westat, the contractor that collects NAEP data for NCES. For the eighth-grade sample, the identification numbers made it possible to link 177 of the 230 school districts included in the NAEP sample. Of the remaining 53 school districts, 5 were linked through address information; the other 48 are private schools and therefore were not included in the CCD. The same procedure was then applied to the fourth-grade sample. Here, it was possible to link 195 of the 270 school districts using identification numbers. Of the remaining 75 school districts, eight were linked through address information. For the resulting data bases, TCI scores for the state of each district were entered manually. In terms of weighting and other issues, the resulting data base took on the sampling characteristics of NAEP, because NAEP was the only database for which cases were a sample rather than the national universe.

The new data bases were used to produce measures of the 10 variables needed to test the hypotheses (detailed at the end of the Appendix). The TCI was used as taken

BEST COPY AVAILABLE

from the NCES study. Per pupil expenditures in the four areas (instruction, central office administration, school-level administration, and capital outlays) were calculated by taking total expenditures in each area from CCD and dividing by the number of students in the district. Teacher-student ratios were calculated by dividing the total number of teachers in the NAEP schools by the total number of students according to the NAEP school questionnaire. The highest degree of the teachers was taken directly from the NAEP question on that subject in the NAEP teacher questionnaire.

Two other variables were constructed from summated scales. Socio-economic status was calculated from a summated scale of seven items in the NAEP student background questionnaire that pertain to the educational levels and economic resources of the students' families (for each individual respondent to the NAEP, whether or not the family receives newspaper; whether or not there is an encyclopedia in the home; whether or not there are more than 25 books in the home; whether or not the family subscribes to magazines; the highest level of education attained by the mother; the highest level of education attained by the father; and for each school in NAEP, the percentage of students who receive reduced-price or free lunches). The school environment was calculated from a summated scale of seven items in the NAEP school questionnaire (for each school in NAEP, the degree to which teacher absenteeism is not a problem; the degree to which student tardiness is not a problem; the degree to which student absenteeism is not a problem; the degree to which class cutting is not a problem; and the degree to which there is a regard for school property; for each teacher in NAEP, the degree to which teachers have control over instruction; and the degree to which teachers have control over course content). Each plausible value achievement score was taken from the NAEP fourth- and eighth-grade mathematics assessment examinations.

Method

Data were analyzed through the application of a linear structural modeling program known as LISREL 8. The process of analyzing data using LISREL 8 generally involves five steps (Hayduk 1987). First, correlation matrices, means, and standard deviations of the concepts are calculated from the data base. These statistics summarize the characteristics of the "real" data. Second, a model of relationships between variables is specified based on hypotheses concerning those relationships. The hypothesized relationships are not bivariate but, rather, are intended to measure the relationships, known as direct effects, between each of the variables holding the relationships between the others constant. These relationships are indicated through the creation of two matrices, one specifying whether or not relationships are hypothesized between each of the endogenous and exogenous variables, and one specifying whether or not relationships are hypothesized among endogenous variables. (An exogenous variable is one that is hypothesized to influence other variables but is not itself hypothesized to be influenced by any; an endogenous variable is one that is hypothesized to be influenced by at least one of the other variables.) Third, based on the input in the program from the first and second steps, the program generates estimates of the magnitude of these direct effects using the maximum likelihood algorithm. Those variables for which no relationship is hypothesized are treated as if the relationship is zero. Fourth, the goodness-of-fit between the correlation coefficients, means, and standard deviations implied by the hypothesized model and the correlation coefficients, means, and standard deviations from the real data are estimated. Goodness-of-fit typically is measured through a chi-square statistic. Rather than the more typical use of the chi-square statistic, in which a significant chi-square indicates a positive finding, for LISREL an insignificant chi-square indicates a positive finding. The reason for this reversal is that the chi-square measures the degree of difference between the hypothesized model and the real data; if the chi-square is significant, it indicates that there is a large difference between the model and the data (i.e., poor model fit), while if the chi-square is insignificant, it indicates that there is little difference between the model and the data (i.e., good model fit). Goodness-of-fit is sometimes also measured through indexes, known as goodness-of-fit indexes. Those typically used are the adjusted goodness-of-fit index (AGFI) and the normed goodness-of-fit index (NFI). Generally, indexes above .9 are considered to indicate an acceptable goodness-of-fit. Fifth, the model produces indexes, referred to as "modification indexes" that suggest ways in which the goodness-of-fit of models can be improved by changing hypothesized relationships in the model.

Two additional methodological issues about using LISREL to test hypotheses should be noted. First, reliable estimates require that the hypothesized model assumes some relationships to be zero. The

source of this assumption must come from other studies or theories in the literature. If all variables are allowed to be related to one another to some degree, the resulting model will be "misidentified," meaning that there will not be a sufficient number of hypotheses to create the bivariate relationships that the hypotheses imply to match against the real data. This means that LISREL is not simply producing relationships from the data but, rather, is testing the viability of a particular hypothesis. Second, in testing goodness-of-fit, it is possible not only to test whether or not a model fits well, but also to compare its fit to that of other models (referred to as "nested" models), in which some of the relationships are set to zero. For instance, if a model that hypothesizes relationships between two endogenous variables and one exogenous variable produces a model with a chi-square statistic of 15.00 that is statistically significant, and another model that hypothesizes a relationship between one of the endogenous variables and the exogenous variable, but not between either of these and the second endogenous variable, produces a model with a chi-square statistic of 2.00 that is statistically insignificant, then the latter model can be said to possess a better fit. For hypothesis testing, it can be concluded that the second hypothesized model is more likely to be an accurate representation of the data than the first.

For this analysis, a series of LISREL models were run to take into account the methodological issues in the NAEP. First, using the eighth-grade sample only, separate correlation matrices, means, and standard deviations were calculated for each plausible value along with the nine other variables. LISREL models were run on each of these five matrices. A sixth model was generated in which the achievement measure used was the normit. (The normit is the normal deviate corresponding to the student's booklet score as a percentage of the total possible score.) A normit model was used to assess the impact of the plausible values methodology on the LISREL analysis. For each matrix, data were weighted by the product of the mean student-level weight for the district and the number of NAEP students in the district, thus correcting for the unequal probabilities of selection into the NAEP sample. LISREL estimates of direct effects were then produced using Maximum Likelihood Estimation for each matrix, and the parameters and goodnesses-of-fit were compared between models. The comparison indicated that, while there were differences between models, all had goodness-of-fit at the .95-.96 level, and the standardized scores for statistically significant direct effects did not differ significantly. Furthermore, no parameter was statistically significant for one model and not the others.

Since variability between plausible values was not great, the means of the correlation matrices, means, and standard deviations were calculated, resulting in a single set of statistics. LISREL estimates were then generated from these statistics. Proper analysis of sample survey data requires that the effects of the sample design be taken into

account. This was accomplished by applying the jackknife technique to the eighth-grade sample. Correlation matrices were generated for eighth graders where each case was weighted by the product of one of the jackknife weights and the number of cases in the district. Since there were 56 jackknife weights, this resulted in 56 correlation matrices, means, and standard deviations. In principle, 56 separate LISREL analyses could then be conducted, producing estimates for all parameters, which could then be appropriately combined to account for the effects of the sample design. Since the combination of all parameter estimates was computationally expensive, an approximation was adopted using design effects estimated from a subset of the parameter estimates. The parameter estimates for the relationship between mathematics achievement and school environment, highest teacher degree, and teacher-student ratios only were recorded for each of the 56 weights, and their standard deviations calculated. The ratios of these standard errors to the standard errors of the original mean estimate weighted by the student base weight were calculated, producing design effects of 1.18 for school environment, 1.78 for teacher's highest degree, and 1.40 for teacher-student ratio. Based on the jackknife analysis, all subsequent models for both grades were run weighted by the student base weight and measuring achievement with the mean plausible value, but with a design effect of 1.75 (the most conservative of the three values).

Two models were run for each sample using these correlation matrices, means, standard deviations, and number of cases. The first was the same model used for the above analyses. Of the exogenous variables, it permitted the Teacher's Cost Index (TCI) to be related to teacher-student ratios and highest teacher degree but not to school environment or mathematics achievement;[23] instructional per pupil expenditures also to be related to teacher-student ratios and highest teacher degree but not to school environment or mathematics achievement; central office administration also to be related to teacher-student ratios and highest teacher degree but not to school environment or mathematics achievement; school administration per pupil expenditures to be related to school environment but not to the other endogenous variables; capital outlays per pupil expenditures to be related to school environment but not to the other endogenous variables; and socio-economic status to be related to school environment, and mathematics achievement

[23] The TCI was included as a variable in the model rather than as an adjustment to each per pupil expenditure measure so that it would influence only those endogenous variables it would make theoretical sense for it to directly influence (teacher-student ratio and teacher's highest degree). To assess the effect of not including the TCI as an adjustment, the models were re-run excluding the TCI from the model but multiplying each per pupil expenditure measure by it; this modification did not significantly alter the results.

BEST COPY AVAILABLE

but not to teacher-student ratios and highest teacher's degree. Of the endogenous variables, it permitted teacher-student ratios to be related to highest teacher's degree, school environment, and mathematics achievement, but not to itself; highest teacher's degree to be related to school environment and mathematics achievement, but not to itself or reciprocally to teacher-student ratios; school environment to be related to mathematics achievement but not to itself or reciprocally to teacher-student ratios and highest teacher's degree; and mathematics achievement not to be related to itself or reciprocally to any of the other endogenous variables.

Structuring a model in this way permits a test of the four hypotheses. It frees all of the hypothesized relationships for estimation, as well as many relationships about which no hypothesis was specified but that may exist, (such as the relationship between teacher's education and mathematics achievement). Only those relationships that are hypothesized not to exist are fixed at zero. These include the relationship between capital outlays and teacher's education (while there may be an indirect relationship here, it is unlikely that money budgeted for capital construction is spent on teacher salaries) and the relationship between student SES and teacher-student ratios (again, while the two may be indirectly related, it is unlikely that the economic circumstances of students' families causes there to be more teachers). If the relationships fixed at zero turn out in fact to be important relationships according to the data, the hypotheses are therefore not correct, and this will be revealed in low goodness-of-fit indexes. If the specified model accurately represents the data, this will be revealed in high goodness-of-fit indexes.

An alternative model is also specified, in which, in addition to the relationships fixed at zero in the previous model, the four relationships corresponding to the hypotheses are also fixed at zero. Thus, the relationship between instructional per pupil expenditures and teacher-student ratios is fixed at zero, the relationship between central office administration per pupil expenditures and teacher-student ratios is fixed at zero, the relationship between teacher-student ratios and school social environment is fixed at zero, and the relationship between school social environment and mathematics achievement is fixed at zero. The degree to which this more parsimonious model represents a worse fit of the data will be reflected in higher chi-squares. If the chi-square is significantly higher, it may indicate that the parsimonious model does not represent those data as well as the full model. Modification indexes can then specify which relationships are responsible for the significant increase in the chi-square.

The creation of these two models, then, represents a testing of the hypothesis that money matters. If the estimates produced by the first model include statistically

BEST COPY AVAILABLE

significant and substantively large coefficients for the direct effects of instructional and central office per pupil expenditures on teacher-student ratios, of teacher-student ratios on school social environment, and of school social environment on mathematics achievement, the model may support the hypothesis that money matters. Additional support for the notion that money matters can be found if the first model produces a goodness-of-fit index (GFI) above .9 and a statistically insignificant chi-square and if the second model produces a GFI below .9 and a statistically significant chi-square, as well as if the large chi-square of the second model is attributable in large part to the four hypothesized relationships, as measured by modification indexes. On the other hand, if any of these tests is violated, it may be that the data do not support the notion that money matters in the way hypothesized here.

The models that were supported by these tests were the ones included in this report. In addition, controlled comparisons of means were calculated for descriptive purposes. Each sample was split into four subsamples based on the measures of socio-economic status (SES) and the Teacher's Cost Index (TCI): (1) an above-average SES, above-average TCI subsample; (2) an above-average SES, below-average TCI subsample; (3) a below-average SES, above-average TCI subsample; and (4) a below-average SES, below-average TCI subsample. Each subsample was then divided into above- and below-average instructional per pupil expenditure groups, and their mean teacher-student ratios were compared. Each subsample was then divided into above- and below-average teacher-student ratio groups. For fourth graders, their mathematics achievement levels were compared. For eighth graders, their social environment scores were compared, and then the subsamples were divided into above- and below-average social environment subgroups and their achievement compared.

VARIABLE DEFINITIONS

Teacher's Cost Index: Taken from Teacher's Cost Index calculated by NCES (1995). Consists of estimates of the market value of teachers when measures of teacher quality and other characteristics are held constant. Estimates averaged at the state level were included for this analysis.

Instructional Per Pupil Expenditures: Derived from data in the Common Core of Data (CCD) for fiscal year 1992. Calculated by dividing total expenditures on instruction, as defined in CCD, for each school district divided by the number of students in the school district.

Central Office Administration Per Pupil Expenditures: Derived from data in CCD for fiscal year 1992. Calculated by dividing total expenditures on central office administration, as defined in CCD, for each school district divided by the number of students in the school district.

School Administration Per Pupil Expenditures: Derived from data in CCD for fiscal year 1992. Calculated by dividing total expenditures on school-level administration, as defined in CCD, for each school district divided by the number of students in the school district.

Capital Outlays Per Pupil Expenditures: Derived from data in CCD for fiscal year 1992. Calculated by dividing total capital outlays, as defined in CCD, for each school district divided by the number of students in the school district.

Socio-economic Status: Derived from data in National Assessment of Educational Progress (NAEP) in Mathematics for 1992. Calculated as summated scale of the following items: for each individual respondent to NAEP, whether or not family receives newspaper; whether or not there is an encyclopedia in the home; whether or not there are more than 25 books in the home; whether or not the family subscribes to magazines; the highest level of education attained by the mother; the highest level of education attained by the father; and for each school in NAEP, the percentage of students who receive reduced-price or free lunches.

Teacher-Student Ratios: Derived from data in NAEP in Mathematics for 1992. Calculated by dividing total number of teachers in school by total number of students in school.

Highest Degree: Taken from data in NAEP in Mathematics for 1992. Consists of the highest level of education attained by teacher responding to NAEP on behalf of individual student.

School Environment: Derived from data in NAEP in Mathematics for 1992. Calculated as summated scale of the following items: for each school in NAEP, the degree to which teacher absenteeism is not a problem; the degree to which student tardiness is not a problem; the degree to which student absenteeism is not a problem; the degree to which class cutting is not a problem; and the degree to which there is a regard for school property; for each teacher in NAEP, the degree to which teachers have control over instruction; and the degree to which teachers have control over course content. Each item is scored on a scale from 1 to 4.

Mathematics Achievement: Taken from data in NAEP in Mathematics for 1992. Consists of the five plausible values for students responding to NAEP.

REFERENCES

Adams, J. E. (1994). "Spending school reform dollars in Kentucky: Familiar patterns and new programs, but is this reform?" Educational Evaluation and Policy Analysis, *16* (4), 375-390.

Barnabel, J. (1994). "$7,512 per pupil: Where does it go?" The New York Times, (January 23), Section 13, P. 1.

Barro, S. M. (1994). Cost-of-education differentials across the states. Washington, DC: U.S. Department of Education.

Barton, P., Goertz, M., & Coley, R. (1991). The State of Inequality. Princeton, NJ: Educational Testing Service.

Celis, W. (1994a). "A long-running saga over Texas schools." The New York Times, (April 10), Sec. 4A, p. 30.

Celis, W. (1994b.) "Michigan votes for revolution in financing its public schools." The New York Times, (March 17), A1.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of Educational Opportunity. Washington, DC: U.S. Government Printing Office.

Coons, J. E., Clune, W. H., & Sugarman, S. D. (1970). Private Wealth and Public Education. Cambridge, MA: Harvard University Press.

Corcoran, T. & Goertz, M. (1995). "Instructional capacity and high performance standards." Educational Researcher, *24* (9), 27-31.

Cuban, L. (1984). "Transforming the frog into a prince: Effective schools research, policy and practice at the district level." Harvard Educational Review, *54*, 129-151.

Finn, J. D. & Achilles, C. M. (1990). "Answers and questions about class size: A statewide experiment." American Educational Research Journal, *27* (3), 557-577.

Firestone, W. A., Goertz, M. E., Nagle, B., & Smelkinson, M. F. (1994). "Where did the $800 million go? The first year of New Jersey's Quality Education Act." Educational Evaluation and Policy Analysis, *16* (4), 359-373.

Fortune, J. C., & O'Neil, J. S. (1994). "Production function analyses and the study of educational funding equity: A methodological critique." Journal of Education Finance, *20* (Summer), 21-46.

Fuhrman, S. H. (1996). "Issues and data needs in school finance." In William J. Fowler (ed.), Selected Papers in School Finance, (pp. 33-47). Washington, DC: U.S. Department of Education.

Glass, G. V., & Smith, M. L. (1979). "Meta-analysis of research on class size and achievement." Educational Evaluation and Policy Analysis, *1* (1), 2-16.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). Fundamentals of item response theory. London: Sage Publications.

Hanushek, E. A. (1989). "The impact of differential expenditures on school performance." Educational Researcher, *18* (4), 45-65.

Hanushek, E. A. (1994). "Money might matter somewhere: A response to Hedges, Laine, and Greenwald." Educational Researcher, 23 (4), 5-8.

Hanushek, E. A. (1994). Making schools work: Improving performance and controlling costs. Washington, DC: Brookings Institution.

Hayduk, L. A. (1987). Structural equation modeling with LISREL: Essentials and advances. Baltimore, MD: Johns Hopkins University Press.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). "Does money matter? A Meta-Analysis of studies of the effects of differential school inputs on student outcomes." Educational Researcher, *23* (3), 5-14.

Kozol, J. (1991). Savage inequalities. New York, NY: Crown

Johnson, E. (1989). "Considerations and techniques for the analysis of NAEP data." Journal of Educational Statistics, *14* (4), 303-334.

Johnson, E. (1994). "Overview of Part I: The design and implementation of the 1992 NAEP." In Eugene G. Johnson & James E. Carlson (Eds.), The NAEP 1992 technical report, (pp. 9-32). Princeton, NJ: Educational Testing Service.

Lee, V. E., Bryk, A. S., & Smith, J.B. (1993). "The organization of effective secondary schools." Review of Research in Education, *19*, 171-267.

Miles, K. H. (1995). "Freeing resources for improving schools: A case study of teacher allocation in Boston public schools." Educational Evaluation and Policy Analysis, *17* (4), 476-493.

Monk, D. H. (1992). "Educational productivity research: An update and assessment of its role in education finance reform." Educational Evaluation and Policy Analysis, *14*, 307-332.

Mosteller, F., & Moynihan, D. P. (1972). "A pathbreaking report: Further studies of the Coleman Report." In Frederick Mosteller & Daniel P. Moynihan (Eds.), On equality of educational opportunity (pp. 3-66). New York: Random House.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1993). NAEP 1992 mathematics report card for the nation and the states. Princeton, NJ: Educational Testing Service.

Murnane, R. J., & Levy, F. (1995). "Evidence from fifteen schools in Austin, Texas." In Gary Burtless (Ed.), Does money matter? The effect of school resources on student achievement and adult success (pp. 93-97). Washington, DC: Brookings Institution Press.

National Center for Education Statistics (1995). Public school teacher cost differences across the United States. Washington, DC: Government Printing Office.

New York Times (1994). "Paying for New Jersey's schools: Excerpts from court ruling ordering parity in New Jersey school funding." The New York Times, (July 13): B7.

Odden, A. (1990). "Class size and student achievement: Research-based policy alternatives." Educational Evaluation and Policy Analysis, *12* (2), 213-227.

Odden, A. (1993). "School finance reform in Kentucky, New Jersey and Texas." Journal of Education Finance, *18* (Spring), 293-317.

Odden, A., & Clune, W. H. (1995). "Improving educational productivity and school finance." Educational Researcher, *24* (9), 6-10, 22.

Petersen, M. (1996). "A bitter taste in Newark schools." The New York Times (September18), B1, B5.

Picus, L. O. (1994). "The local impact of school finance reform in four Texas school districts." Educational Evaluation and Policy Analysis, *16* (4), 391-404.

Purkey, S. C., & Smith, M. S. (1983). "Effective schools: A review." Elementary School Journal, *83*, 427-452.

U.S. General Accounting Office. (1995). School finance: Trends in U.S. spending. Washington, DC: U.S. GAO.

Wenglinsky, H. (1996a). Home, school, and family: The relative influence of family norms and private school control on student achievement. Unpublished Dissertation: New York University.

Wenglinsky, H. (1996b). Modeling the relationship between school district spending and academic achievement: A multivariate analysis of the 1992 National Assessment of Educational Progress and the Common Core of Data. Princeton, NJ: Educational Testing Service.

04202-14510 • S47M6 • 204891 • Printed in U.S.A.